DOCUMENT RESUME

ED 299 264 TM 011 197

AUTHOR Davis, Todd; And Others
TITLE Speededness of the Academic Assessment Placement Program (AAPP) Reading Comprehension Test.
SPONS AGENCY Tennessee State Board of Regents, Nashville.
PUB DATE 9 Jul 87
NOTE 57p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS Admission Criteria; College Admission; College Students; Community Colleges; Higher Education; *Reading Comprehension; Reading Tests; *State Programs; *Student Placement; Technical Institutes; Test Construction; *Testing Programs; Test Wiseness; *Timed Tests; Universities
IDENTIFIERS *Academic Assessment Placement Program; ACT Assessment; Response Patterns; *Speededness (Tests)

ABSTRACT

Whether the time limits of the Academic Assessment Placement Program (AAPP) Reading Comprehension Test of the State Board of Regents (SBR) of Tennessee provide fair treatment for all underprepared students was assessed. Focus was on studying various response patterns and completion rates that indicate whether or not speededness (SP) is critical. Definitions and criteria for test SP and five methods of measuring SP are reviewed. Results of the AAPP testing were studied by age, sex, race, and preparedness as measured by scores on the American College Tests (ACT). The relationship between SP and ability was determined by comparing the completion rates of students with ACT scores above and below 16. Completion rate data for the SBR system allowed comparison of students from four-year institutions, community colleges, and state technical institutions. For the total population of 8,290 taking the test, no substantive indications of SP were found for males or females. Students aged over 21 years had lower completion rates than did younger counterparts. Of the 4,946 white students, 88.9% completed the AAPP, while of the 1,954 black students, only 73.7% completed the AAPP. This was generally true for all campuses in the system. An additional 10 minutes was recommended to allow greater completion of test items. Twenty-six graphs illustrate test results for the different student populations. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED299264

SPEEDEDNESS OF THE ACADEMIC ASSESSMENT PLACEMENT PROGRAM (AAPP) READING COMPREHENSION TEST

A Report Submitted to the Board of Regents of the State University and Community College System of Tennessee

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

TODD DAVIS

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

July 9, 1987

Todd Davis
Robert Kaiser
Jerry Boone

Center for Higher Education
MEMPHIS STATE UNIVERSITY

BEST COPY AVAILABLE

TM 011 197

Speededness of the Academic Assessment Placement Program (AAPP) Reading Comprehension Test

Background Concerns

Underprepared students are applying to State Board of Regents (SBR) institutions in large numbers. In the fall of 1986, (Report 1) 9,588 first-time freshmen under age 21 applied to SBR institutions and 5,254, or 54.8 percent, were enrolled in one or more remedial or developmental courses.

Students under 21 years of age scoring 15 or below on their American College Tests (ACT) are required to take the Academic Assessment and Placement Program (AAPP) tests in English, mathematics, reading comprehension and logical relationships. Students 21 years of age or older, having no ACT test on file, are also required to take the AAPP tests. Performance on these tests determines if students are placed in non-credit-producing remedial or developmental courses. These courses are designed to establish the additional background necessary for successful completion of college level coursework. If students score highly enough on the AAPP, they are permitted to enroll in regular credit-producing college courses. Specifically

related to reading, 1,508 or 15.7 percent of the first-time freshmen in fall, 1986 did not score highly enough for collee credit placement and were enrolled in remedial reading courses while 2,107 or 21.9 percent were enrolled in developmental reading courses (Report 1).

The consequence of a poor performance on the AAPP is that students spend additional time (semesters) and money (tuition) in their quest for a college degree. In addition, the number of underprepared students identified by the ACT reinforces the need for a high level of confidence in the use of the AAPP. This high level of confidence is an important consideration because it has state-wide programmatic and financial implications.

This present investigation was motivated by the question: Do the time limits of the reading comprehension section of the AAPP provide fair treatment for all underprepared students? This is a significant question because test publishers regard this type of test (timed-power) as an appropriate test for student placement. Speededness, or the imposition of time limits, is the characteristic of the AAPP test which caused the most concern for instructors and directors

when the test was administered for remedial and developmental student placement. Consideration of this factor initiated the SBR investigation and this report is the product of a team of investigators from Memphis State University who were assigned the task of researching the question: **Do the time limits imposed by the AAPP foster differences in the performance of students with respect to their ability, age, sex, race and institution?**

To verify the concern for test speededness a questionnaire was sent to each of the 20 directors of developmental programs in the SBR system. They were asked to give an open-ended expression of concerns they, their instructors, or their students had with the time limits on the AAPP reading comprehension test. Seventeen directors responded to the inquiry and a few (6) had no concerns with the time limits. In what appeared to be an informal manner, several directors surveyed test results from their institutions and concluded that, given more time, many students would have done better. They suggested that a 45 minute time limit would have been more equitable than the 30 minute limit which has been used.

Specifically, some of the directors felt that the AAPP reading comprehension test was designed for all students entering college and not just those suspected of being underprepared. One director commented that students typically should be expected to answer two-thirds of the questions correctly (30 of 45) within the time allotted. Under the current SBR requirements, a score of 36 is necessary for a student to bypass developmental courses. The director felt that this was an unrealistic expectation. If this director is right, most students would be held to remedial or developmental courses by the present time limit.

There was a concern that the time restrictions did not represent developmental studies goals and objectives found in the "Green Book" (Academic Preparation for College: What Students Need to Know and Be Able to Do) (1983). Instructors felt that the test, with its short reading passages, did not represent realistic expectations of the time and tasks required to complete college homework and class assignments.

Another concern was related to anxiety and time limits. Some students apparently became anxious and were encouraged to guess when they neared the end of the time period. They were reported to have used this as an

excuse for poor performance when confronted with their test results. Others directly stated that time restrictions inhibited concentration and caused them to think that they had to hurry and attention to detail was lost.

"Background knowledge" was cited as a concern related to time restrictions, although the comment was not clarified by the director who mentioned it. (It was included with a comment related to the notion that the AAPP, "does not allow the student to clarify thinking processes. . .".) Cultural differences may have been at the root of the director's comments, but this was not clear from the short response.

Our questionnaire to directors asked only about the placement test, Form A. It did not address the use of Form B as a posttest. Yet some of the respondents used the opportunity to question the use of the posttest as a possible measure of program effectiveness so long as it is given under quite different conditions than those in force at the time of the pretest.

In summary, the directors and instructors expressed a number of different concerns about the use of the AAPP with time limits for student placement in remedial and developmental reading courses. Their concerns were

expressed in a genuine, non-critical fashion and served to underscore the importance of this study.

Organization of This Report

The initial sections of the report are designed to give background information for the SBR-sponsored project on the speededness of the reading comprehension test. First there is information on the approaches used to establish time limits. Next there is a section which reviews various definitions and criteria for test speededness, allowing the reader to understand the criteria used to judge the speededness of the AAPP in the present investigation. A section on earlier research relating speededness to age, race and sex appears next. Concluding the introductory material is a section describing the design of the present study.

The results section of the report has been organized to present information on speededness using completion rates for the AAPP test of reading comprehension. Graphs that plot acceptable and unacceptable standards of timed-power test speededness are used to illustrate the SBR data for the fall of 1986. These graphs also reveal the speededness characteristics of the AAPP as they relate to ability, age, sex, race and institution.

Time Limits

From a survey of professional literature, three kinds of tests have been identified which bear on the present question, speed tests, power tests, and a combination of the two. Speed tests are constructed with all items at one level of difficulty and, when time limits are imposed, the test measures a characteristic involving a subject's accuracy and rate of work. Conversely, power tests measure performance when no time limits are imposed. Generally, these tests grow progressively more difficult, item by item, and subjects determine their own time of completion.

A key factor in the selection of one kind of test or another is the ultimate use of the test. If the test is used to predict performance where speed is a significant factor, then a timed test should be employed. If speed plays no important role in the prediction of a criterion variable, then a power test is more appropriate.

A third kind of test has emerged that represents a compromise between speed and power tests. This is a power test that has a time limit imposed on it, but the limit is imposed in such a way that it should have minimal influence on a typical subject's score. In such

a test, item difficulty increases and a specified percentage of subjects (80 or 90) are expected to finish the test.

Definitions of Speededness

The measurement of test speededness has been the subject of many investigations. Five methods of evaluation speededness have evolved, and with them, different operational definitions. Donlon (1980b, p. 4) identifies and discusses the strengths and weaknesses of each in a technical article that explores the implications of test speededness. Briefly these approaches are:

Gulliksen This approach to measuring speededness is based on two ratios: (1) the ratio of unattempted items to items answered wrong and, (2) the ratio of unattempted items to total score. This method does not distinguish between items Not Reached and items omitted; both are items unattempted.

Cronbach & Warrington This approach is based on the correlation of scores obtained with a time limit and scores obtained without a time limit. This method requires that the subjects take equivalent forms of the test, one timed and one untimed.

Stafford This approach computes a measurement of speed by analyzing unattempted items. It does not distinguish between items omitted and items Not Reached.

ETS (Educational Testing Service) This approach is based on the number of items completed by those who take the test. ETS considers a test to be speeded if fewer than 100 percent of the candidates reach 75 percent of the items and fewer than 80 percent of the candidates finish the test.

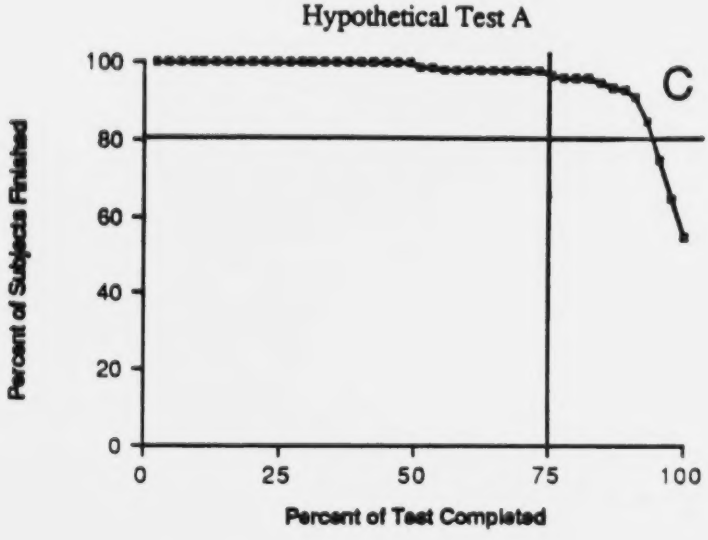
Reilly-Donlon This approach to measuring speededness of a test is based on the proportion of subjects who do not finish the test in the allotted time and on the number of items Not Reached, as opposed to the number of unattempted items. The approach is derived from the ETS analysis and uses a Not Reached score as the fundamental measure of speed. The number of items Not Reached is defined by a continuous string of items running from the last item completed to the end of the test. For example, in a 45 item test, such as the AAPP, if a subject fails to

respond to items 40, 43, 44, and 45, the number of unattempted items would be four (4), while the number of Not Reached items would be three (3).

Donlon (1980b, p. 40) discussed these measures and demonstrated that Stafford's approach is basically like Gulliksen's in that the measurement of speed is considered as a class of items called "unattempted." He points out that Stafford does not make a distinction between items omitted and items Not Reached. This is a weakness in some instances, because an omitted item may take as much time (or more) as an item that has a response. People with the same Not Reached score, but different omit scores, will not have performed at the same rate of work. Thus, the people who take time to read and work out an answer, but choose not to respond, do not complete as many items as those who systematically work more rapidly. Donlon (1980b, p. 42) reasons that only when there is some correlation between the rate of success and the rate of attempts, is there evidence that the standing of individuals might not be changed with more time. If omissions are not eliminated from the analysis of unattempted items, then (in many instances) the scores of subjects will be likely to change.

In an earlier article, Donlon (1975, p. 5) describes a graphic technique used by Evans and Reilly to plot ETS criteria for speededness. This diagram pictures the criteria in terms of percent of the subjects still working at the end of a timed test and the percent of test items completed.

Figure 1



The dotted lines in Figure 1 indicate that for a test to be unspeeded, all of the subjects should have completed 75 percent (or better) of the test and 80 percent of the subjects should have completed all of the test. Quadrant C represents the area into which a test's response pattern must fall if it is to meet the

ETS criteria for an unspeeded test. In this area a portion of the subjects are still working toward completion of the test, but not more than 20 percent fail to complete all of the items.

The graph is a plot of Not Reached items expressed as a percent. Item difficulty for individual items on a test can be seen in a steep slope on the graph. A hypothetical test has been plotted to illustrate a test that is basically unspeeded by ETS standards. The graph shows items that precipitate a steep slope especially near the end of the test. The steep slope indicates that subjects have been prevented from making an attempt on the last items. A Not Reached pattern of this sort should be avoided in the development of a power test or a timed-power test that is free from the influence of a time limit. The Multiple Assessment and Placement Services (MAPS) instrument, from which the AAPP is derived, is such a timed-power test.

Review of the Literature

Previous studies of speededness have examined the variables of ability, age, sex and race, but few specific conclusions and generalizations can be drawn. Investigations of ability have focused on a variety of different topics including foreign languages, mental

abilities or achievement test batteries. Age research has dealt with studies on adults and on elementary grade levels. Research designs concerning sex have combined it with other variables like race and have focused attention on Native Americans, blacks and whites.

Several of these studies have been reviewed to illustrate the notion that the question of speededness has not yet been resolved and more research on this topic is necessary.

There is no shortage of research on speededness. It has been the subject of many investigations and can be found by surveying the literature under the topic of "timed tests." For years, investigators have expressed concern about imposing time limits on tests of one kind or another. These studies contain information that is both interesting and relevant to those charged with the responsibility of establishing a testing program.

In Florida, Gallagher (1985) investigated the time limits of the College Level Academic Skills Test (CLAST) to see if it was biased against non-native English speakers. When additional time limits were allowed for subjects in two sample groups, no significant relationships could be found between race/ethnic characteristics and extra time. However, a low failure

rate limited the author's ability to define the relationships among variables.

Jolly and others (1985) investigated the effect of test speededness on random guessing using the scores from 50,000 subjects who had taken Stanford Achievement Tests in Palm Beach, Florida. Inconsistencies between the reading comprehension test scores and other subtest scores had been noted, especially in grades four through seven. Students who finished the reading comprehension subtest were not the achievers on other subtests. Apparently the speededness of the reading subtest encouraged random guessing. New norms were developed for a version of the test with fewer items in order to overcome the validity problems of the longer version.

Flores and Seaman (1978) conducted a comparative study with adults in Texas who were given timed and non-timed versions of the General Educational Development tests (GED). Each of two groups had 300 students drawn from 26 different testing centers. Analysis of the data indicated that the element of time did not make a difference in student performance. Individuals who performed better on non-timed tests included: individuals under 20 years of age and those over 39 years of age, males, persons who had completed 10 or

more years of schooling, black Americans and students who had participated in GED preparatory classes.

Immerman (1980) used the blue level of the Stanford Diagnostic Reading Test (SDRT) to study the effects of eliminating standardized test time restraints on American Indian adults. His results indicated that the experimental group with no time restraints scored higher than a control group with time restraints. One of his interesting recommendations concerned the elimination of "time stress" which is a condition avoided in many Indian cultures.

Reilly and Evans (1974) noted significant (cultural) group-by-time interactions in early research on speededness. They conducted a study to test the effects of time limits on culturally defined groups. They extended the time limits of a national aptitude test from 30 to 40 minutes for a group of college seniors. They rejected the notion that more liberal time limits would benefit minority groups.

There are many other studies available for review. Donlon (1980a) has prepared an annotated bibliography that summarizes the research from 1902 to 1976. There is ample information in this review to initiate the study of any aspect of speededness research. However,

there has not been a clear and concise summary that provides a practitioner with specific recommendations. The process of abstraction from research is difficult and, until such time that these conclusions are available, a test program should include its own research if it is to be carried out with confidence.

Design of the Study

This study design called for the investigation of various response patterns and completion rates that would indicate whether or not speededness is critical. An analysis of completion rate data for the SBR system and individual SBR institutions allowed for comparison of student populations in three types of institutions: (a) the four year institutions, (b) the community colleges and (c) the state technical institutes.

The relationship of speededness and ability was addressed with an analysis of response patterns and ACT performance. Students who have scored lower than 16 on the ACT were the focus of interest in this study. Response patterns within this homogeneous group and a comparison of completion rates between this group and a smaller group who have scored above 16 illustrate the relationship of speededness and ability.

The relationship of speededness and age was addressed with two populations of students; one group 21 years of age and older and another group under 21 years of age. Comparison of response patterns between these groups indicate whether older students respond more or less favorably than younger students to testing situations that require time limits.

Race, sex, and institutions were also considered as a part of this study. A variety of interactions are available for review. An SBR system profile concerning the influence of speededness on this reading comprehension test has become evident in this report.

Data for this project were analyzed on the basis of Donlon's (1980b, p. 52) suggestions about graphing test results with attention focused on items Not Reached, ignoring omissions. An additional reason for using this approach was the consistent relationship it had with the definition of speededness used by ETS, the developers of the AAPP. Like ETS, the results of this study have been plotted on a graph. Unlike ETS, the data plotted on the graph were the percent of the group finishing a test and the item numbers of the reading comprehension test.

This change constitutes a slight modification of the ETS graphic presentation. The results are reported for ability, age, sex, race and institution.

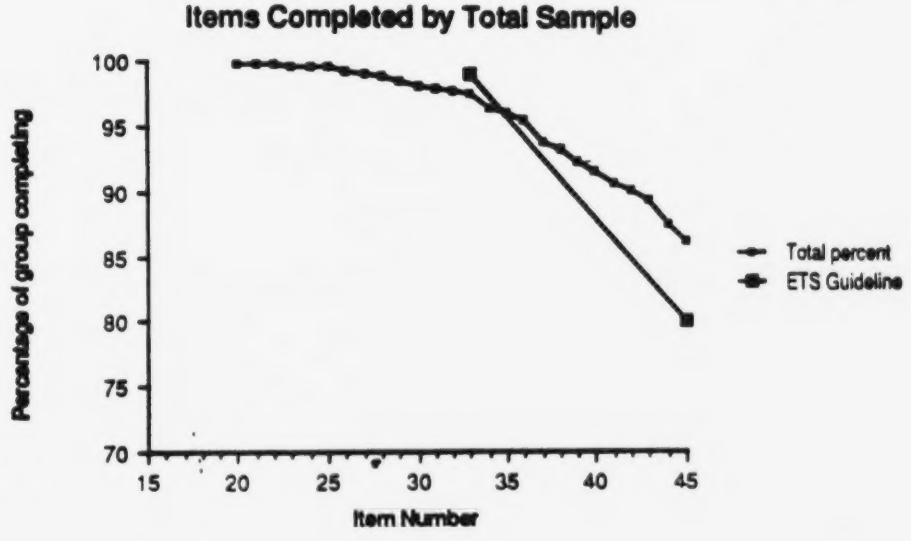
Results

The results of this study on the speededness of the AAPP reading comprehension subtest are graphically presented for the entire population of students who took the test during the fall of 1986 and separately by ACT score, sex, age, race and by racial groups within the 20 SBR member institutions. The question of the speededness of the AAPP is addressed through the examinations of curves obtained by plotting the percent of examinees who have reached the last 30 items on the 45 item AAPP reading comprehension subtest. The ETS criteria for a timed-power test are also plotted on the graph. This graphic display technique follows that suggested by Evans and Reilly and discussed previously in this report. Specifically, the criteria against which the AAPP is compared are: (1) virtually all test takers must reach item 33 (75 percent of the items) and, (2) 80 percent or more of the test takers must reach all of the items. If both of these criteria are met, the AAPP may be judged to be essentially unspeeded.

Speededness for the SBR System

Inspection of Figure 2 suggests that for the total population of SBR students who took the AAPP reading comprehension subtest, no substantive indications of speededness are noted. This plot is based on the responses of 8,290 students. The mean number of items Not Reached for this total group is 1.129 with a standard deviation (SD) of 3.688.

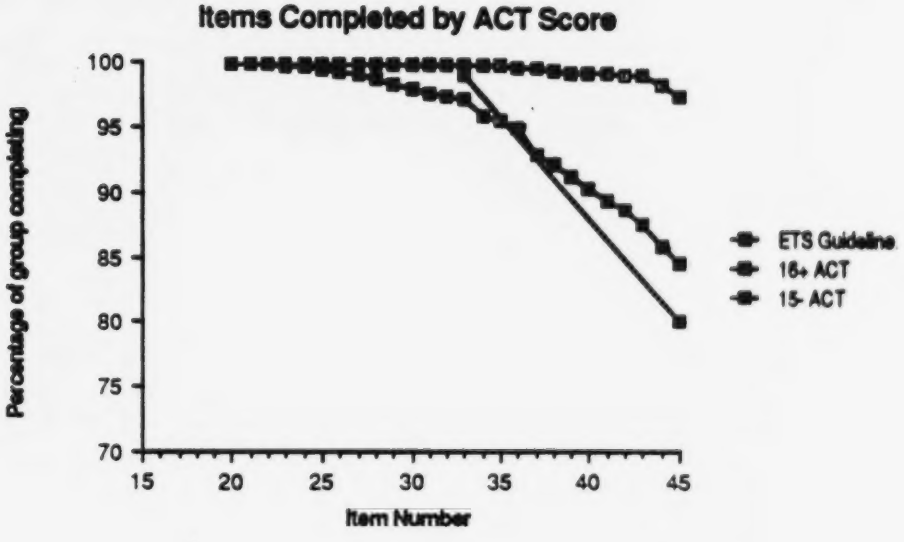
Figure 2



It is the policy of some of the SBR member institutions to test all incoming students with the AAPP regardless of their scores on the ACT. This policy made

it possible to compare the relative speededness of the AAPP for those with passing ACT Composite scores of 16 or greater with those who obtained ACT Composite scores of 15 and less. Figure 3 presents the comparison of these two groups with respect to the speededness of the AAPP.

Figure 3



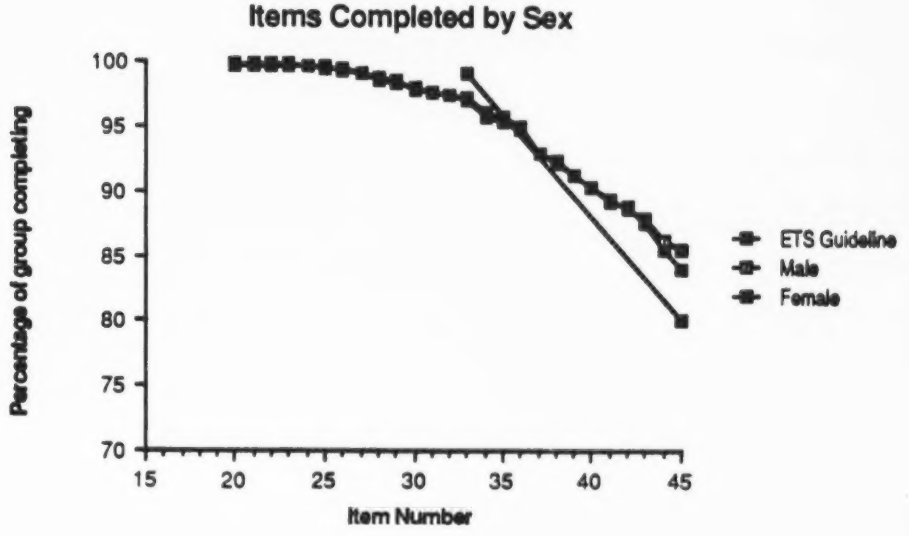
Of those taking the AAPP reading comprehension subtest, 7,088 or 85.5 percent earned a score of 15 or less on the ACT, while 1,202 or 14.5 percent earned a score of 16 or greater. As Figure 3 clearly indicates, a substantial difference between these two groups can be

noted. For the higher ACT group, virtually all of the respondents completed the entire test. The mean number of Not Reached items for this group was 0.131 with an SD of 1.518 items. Inspection of the curve obtained for the lower ACT group indicates that, while speed appears to be somewhat more important for this group, it conforms to the ETS guidelines with 84.5 percent of respondents completing the test, and all but 2.7 percent completed at least 75 percent of the test items. The mean number of Not Reached items for this lower ACT group was 1.298 with a SD of 3.914. The difference between the two ACT groups appeared to suggest that speededness was not a factor in the performance of the higher ACT group, and these 1,202 high scoring individuals, who presumably were ineligible for either remedial or developmental placement, were excluded from subsequent analyses.

Speededness by major classifications. The records of the 7,088 students who scored 15 or less on the ACT were analyzed separately by sex. Of this group, 4,171, or 58.8 percent, were females; 2,848, or 40.1 percent, were males and the sex of 69 individuals was not identified. Figure 4 displays a comparison of the

speededness curves developed for males and females who scored less than 16 on the ACT.

Figure 4

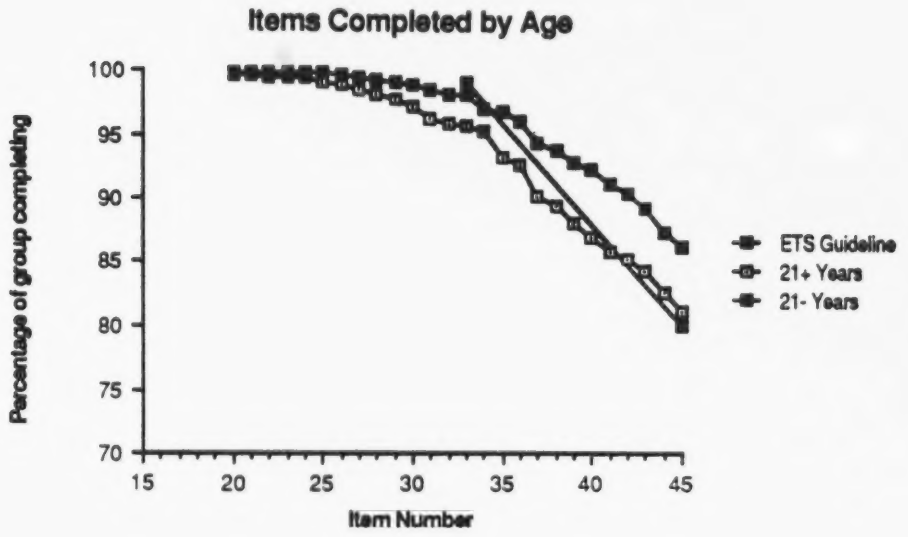


As the graph indicates, speed does not appear to be an important factor for either males or females and virtually no difference in the performance of the two sexes is noted. The mean number of Not Reached items by males and females was 1.296 and 1.287, respectively, with a SD of 4.080 and 3.749, respectively.

The relationship of student age and speededness of the AAPP was explored for the low ACT group by splitting the pool of test takers into two groups and plotting

their completion curves separately. The older group consisted of 2,423 (34.1 percent) students who were 21 years of age and older while the larger younger group consisted of 4,665 (65.8 percent) students who were less than 21 years of age.

Figure 5



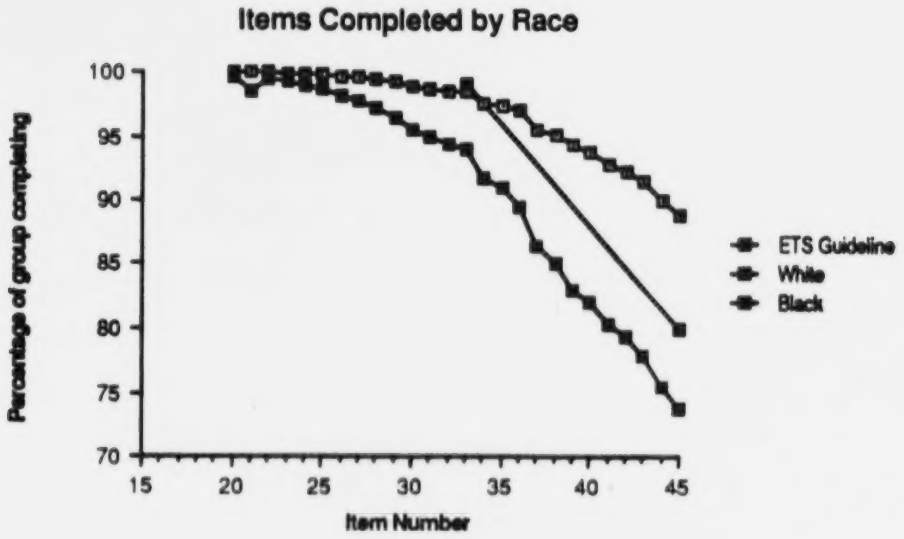
Examination of Figure 5 indicates, that for the younger group, the rates of completion appear to be satisfactory. The older students, while marginally within the ETS guidelines for a non-speeded test, appear to have a markedly lower completion rate than their younger counterparts. Among the younger group, 86.2

percent reach the last item while this is true of only 81.1 percent of older students. On the average the younger group fails to reach 1.03 items (SD=3.34) and the older group fails to reach 1.79 items (SD=4.78). As a generality neither group can be said to be affected by the time limits of the AAPP. However, there are proportionately more older students among the low ACT group who do appear to have difficulty completing the test within the specified time limit.

It was expected at the outset of this study that the AAPP could be examined for possible speededness for each racial group represented within the SBR system. When the pool of low ACT students was split by race, 47 reported themselves as Asian, 33 as Native American, and 39 as Hispanic. Black students (n=1,954) represented 27.5 percent of the low ACT group and whites accounted for 69.7 percent with 4,946 students. As there were so few Asians, native Americans, and Hispanics in this population, fewer than 0.02 percent in total, it was felt that a reliable estimate of performance on the AAPP with respect to speed could not be obtained for these groups. Therefore, reliably stable comparisons by race may only be made for blacks and whites. Figure 6

displays a striking difference between blacks and whites with respect to their ability to complete the AAPP within the established time limits.

Figure 6



While 88.9 percent of white students are able to complete the AAPP and 98.5 percent, or virtually all, white students reach at least 75 percent of the AAPP items, only 73.7 percent of blacks are completing the test with only 94.4 percent completing at least 75 percent of test items. On the average, blacks fail to reach 2.4 items (SD=5.29) while whites fail to reach .82 items (SD=2.97) of the 45 items on the AAPP reading comprehension subtest. This graph indicates a marked

difference in completion rate between blacks and whites. Using the ETS standard referred to earlier, speededness was an important factor for blacks who took the test, but it was not for whites.

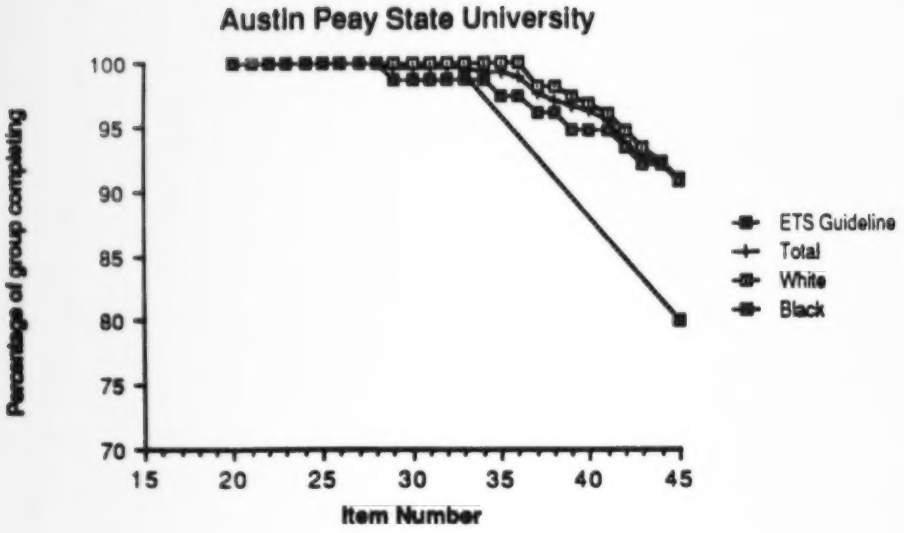
Speededness by institution and race. The SBR system is composed of 20 institutions. Six of these units are universities. These "senior" institutions have historically served a related but separate mission with respect to the ten SBR community colleges. The "junior" institutions, in turn, serve a different population and perform a different mission from that of the four SBR technical institutes. As these three types of institutions are dissimilar in many important ways, the effect of speededness on students' testing at these schools in best gauged comparatively within institutional type.

Race appeared to be a major factor in understanding the implications of the possible speededness of the AAPP on student performance. Therefore, a breakdown of student populations at the institutions by race was performed. Each institutional graph depicts the performance of the total population of low ACT

performers at that institution as well as the performance of blacks and whites separately. By comparing the race plots within particular schools and within the different institution types, a more complete understanding of the differential performance of whites and blacks as groups is obtained.

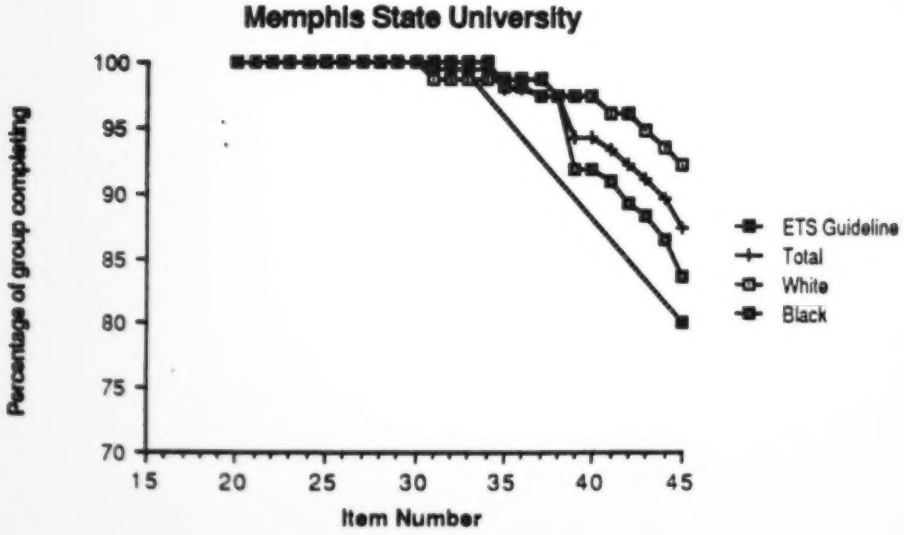
Speededness for the senior institutions. The senior SBR institutions include East Tennessee State University, Memphis State University, Middle Tennessee State University, Tennessee Technological University, Tennessee State University, and Austin Peay State University. The records of 2,291 students attempting to enroll in these senior institutions were examined. The AAPP appears to function as a timed-power test for these institutions taken as a whole. When the completion rates for these schools are examined for possible differential completion rates by race, white students appear to have higher completion rates than black students in all of these schools. Examination of the plots for Austin Peay (Figure 7), Memphis State (Figure 8), Middle Tennessee State (Figure 9), and Tennessee State (Figure 10) indicate that for both whites and blacks, the AAPP is functioning as a timed-power test and meets the ETS criteria for a non-speeded test.

Figure 7



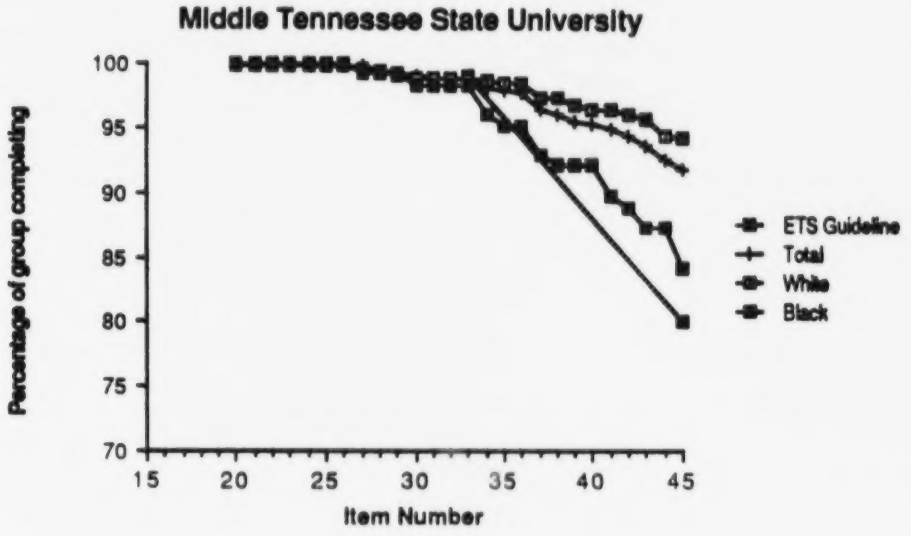
At Austin Peay a negligible difference in the completion rates of white (n=155) and black (n=76) students is noted. Over 90 percent of whites and blacks are able to complete the AAPP within the allotted time.

Figure 8



At Memphis State the records of 110 blacks and 76 white students are plotted. For both groups, the ETS criteria for a timed-power test are sustained, even though there is a prominent difference in completion rates for white and black students.

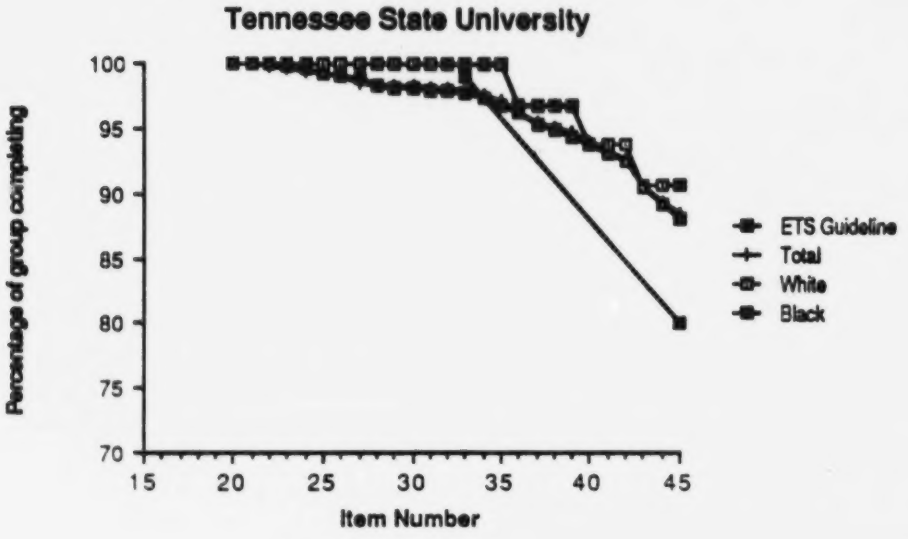
Figure 9



At Middle Tennessee State the performance of 126 black students and 462 white students is graphically presented. At Middle Tennessee, as with Memphis State, the AAPP functions as a timed-power test for both blacks and whites. Even though the completion rates of whites and blacks are quite disparate, both white and black

performance are within the ETS guidelines for a timed-power test. Among white students, 94.2 percent complete the AAPP within the 45 minute time limit as compared with 84.1 percent of blacks. For both whites and blacks, virtually all students are able to reach at least 75 percent of the items.

Figure 10

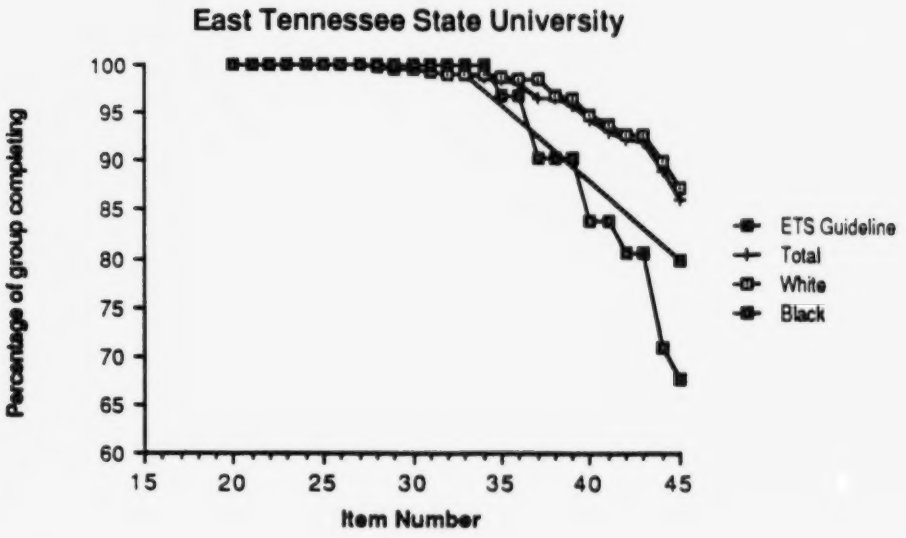


At Tennessee State, 451 black students were tested, as were 32 whites. As the graph indicates, only negligible differences between the races are indicated and for both groups the AAPP functions as a timed-power test. The jaggedness of the curve for white students is

a reflection of the small number of observations. For example, the jump between items 35 and 36 actually reflects the completion of just one student although, by percentage, that student represents 3.1 percent of the white students taking the test. By contrast, one black student represents only 0.2 percent of black test takers.

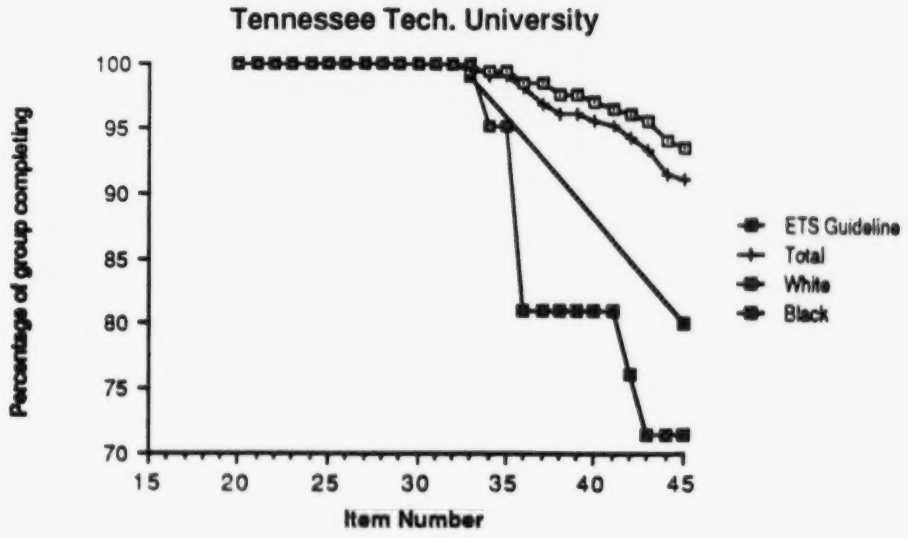
At East Tennessee State and at Tennessee Technological University, major differences in the completion rates of white and black students are noted. These differences are so profound as to suggest that the reading comprehension subtest of the AAPP functions differentially for white and black students at these two schools. Figure 11 displays the results of the analysis of the records of 31 black students and 495 white students at East Tennessee. As the graph indicates, for white students, with a completion rate of 86.6 percent, the AAPP functions as a timed-power test. For blacks, with a completion rate of 67.7 percent, the AAPP contains an important speed component in addition to a comprehension component.

Figure 11



At Tennessee Technological University 21 blacks and 199 whites were tested. As Figure 12 indicates, large differences in the completion rates for whites and blacks are apparent. The jaggedness of the curve for blacks is a reflection of the very small number of blacks tested at this school. Of those who were tested, only 71.4 percent completed the AAPP, as compared with 93.5 percent of white students.

Figure 12



While the number of blacks tested at East Tennessee State and at Tennessee Techological University is small, the differences between their performance and that of whites is apparent. Whether this difference is particular to this present sampling or is a stable one at these schools is unknown.

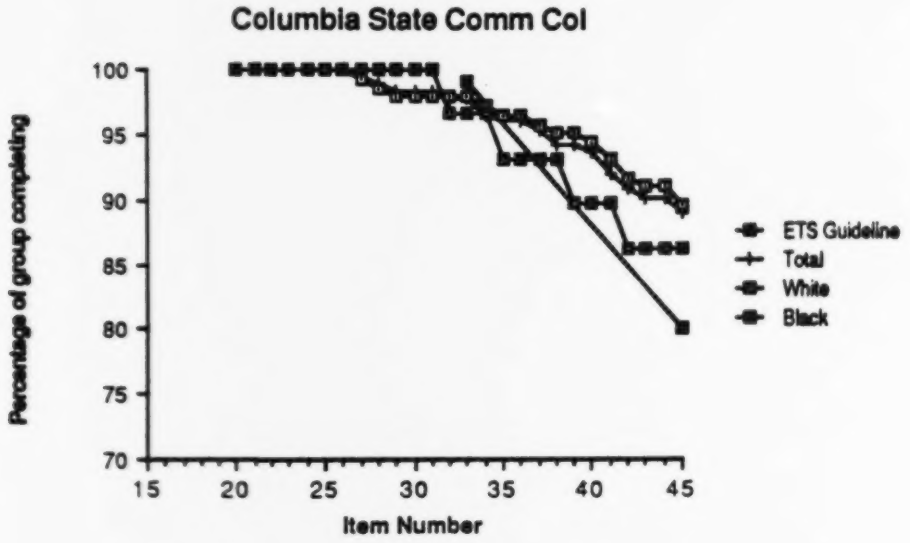
Speededness for the junior institutions. The State Board of Regents operates ten community college throughout the state. These institutions are Columbia State, Jackson State, Cleveland State, Dyersburg State, Motlow State, Walters State, Shelby State, Volunteer State, Roane State and Chattanooga State. Among these

schools the records of 3,151 students were examined. Jackson State is excluded from this analysis because of unaccountable and disproportionate race codings for the school. As this institutional analysis is keyed to potential racial differences, it was felt that Jackson State data should not be included until the apparent anomalies are clarified.

Analysis of the results obtained for blacks and whites at the community colleges indicates differences in the completion rates between blacks and whites at all of the system's community colleges. In all cases, white students' completion rates were higher than those of black students. At seven of the nine schools examined, the differences between blacks and whites were so profound that AAPP performance for blacks, when evaluated by the ETS guidelines, contained an important speed component that was not present for whites. That is, for these black students, both time limits and comprehension were exerting an influence on final scores on this subtest. For whites, the AAPP was predominately a measure of reading comprehension.

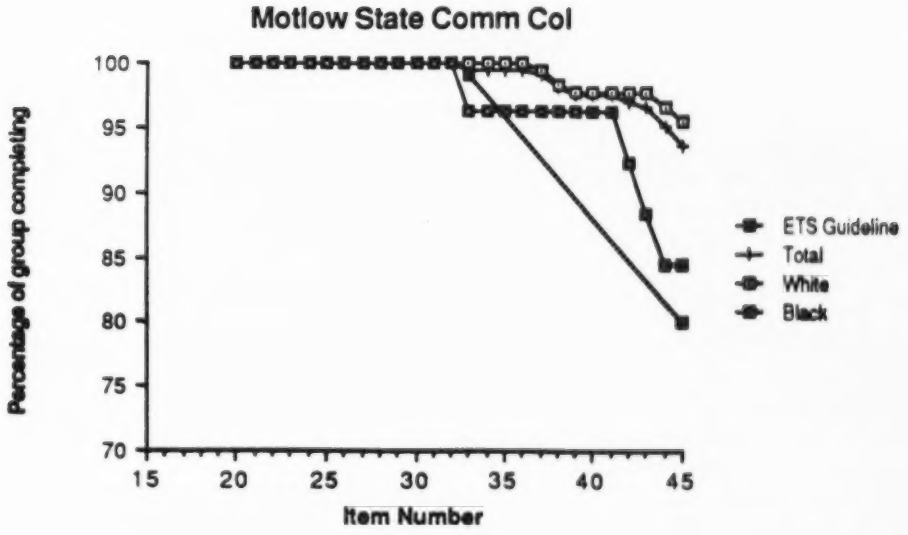
The two schools where the differences between black and white completion rates were the smallest were Columbia State and Motlow State. Figure 13 displays the results for Columbia State.

Figure 13



For Columbia State the graph depicts the performance of 29 black students and 144 white students. Figure 14 displays the results for Motlow State.

Figure 14



At Motlow State 26 blacks and 177 whites were tested. While the completion rates for both groups are within the ETS standards for timed-power tests, a difference in completion rates is apparent. Over 95 percent of white students were able to reach the last item on the AAPP, while only 84.6 percent of blacks were so able.

All of the remaining seven community colleges have marked differential performance between blacks and whites in terms of their completion rates. Figure 15 displays the results for Shelby State, Figure 16 for Volunteer State, Figure 17 Chattanooga State, Figure 18

Roane State, Figure 19 Walters State, Figure 20 Dyersburg State, and Figure 21 Cleveland State.

Figure 15

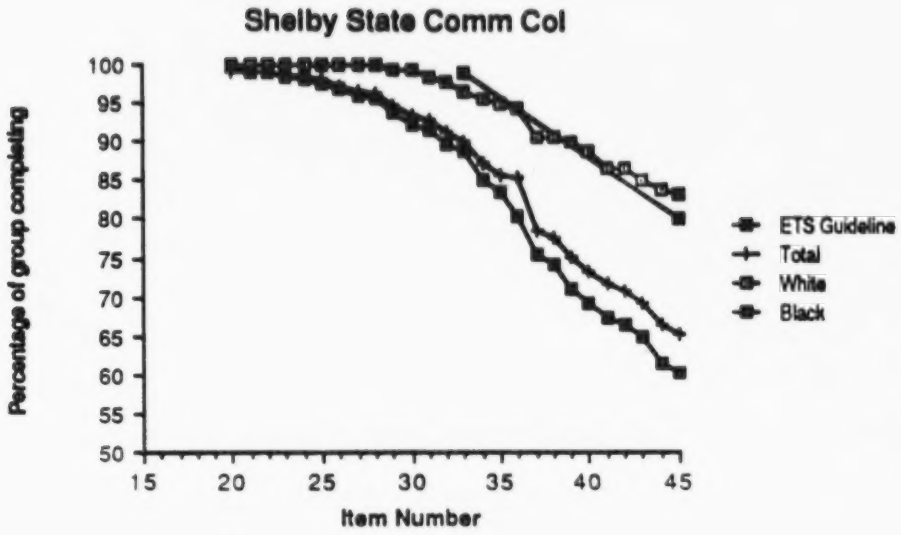


Figure 16

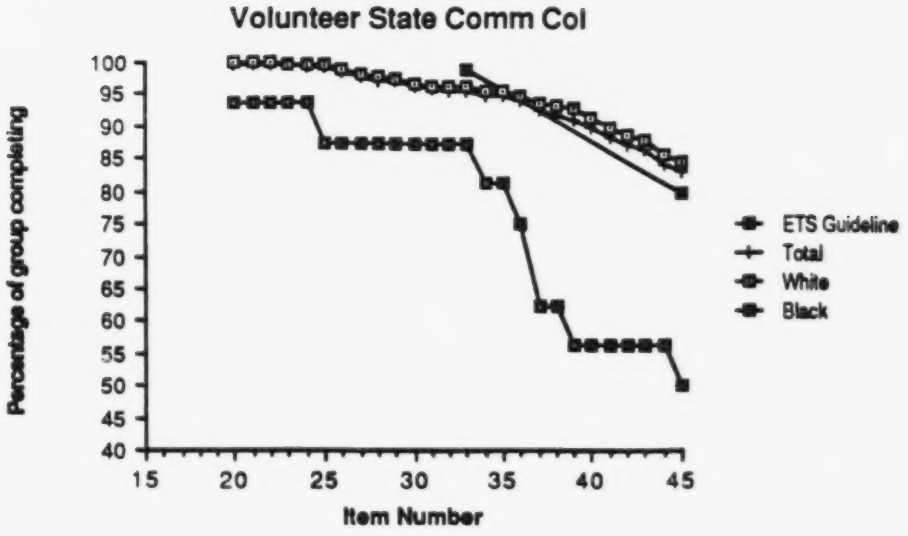


Figure 17

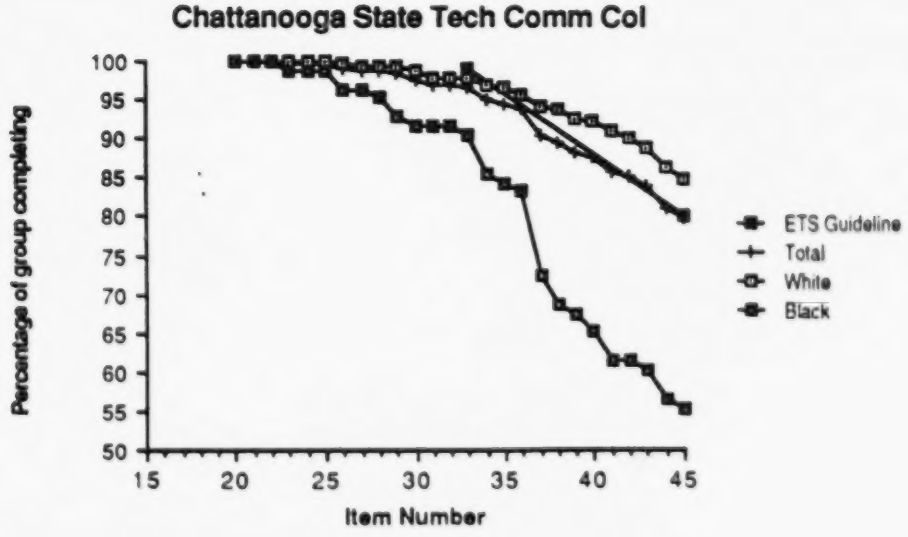


Figure 18

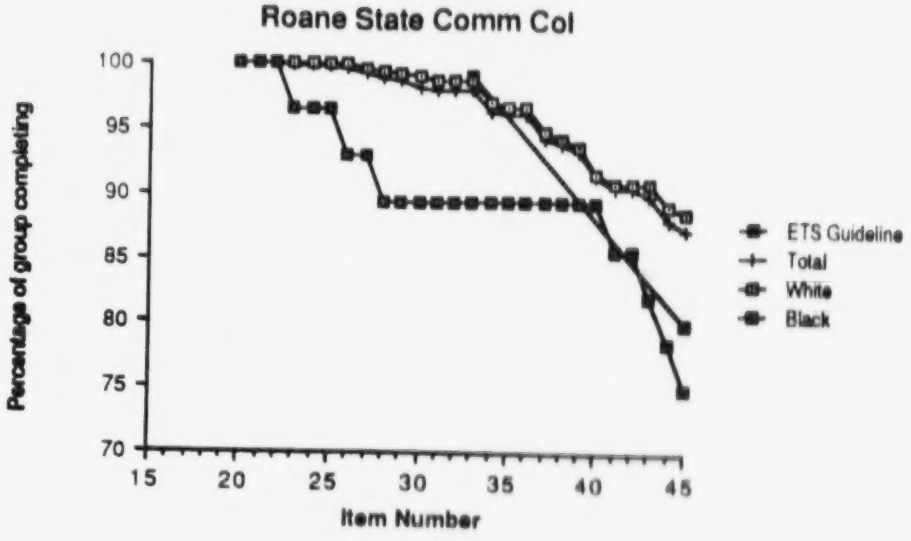


Figure 19

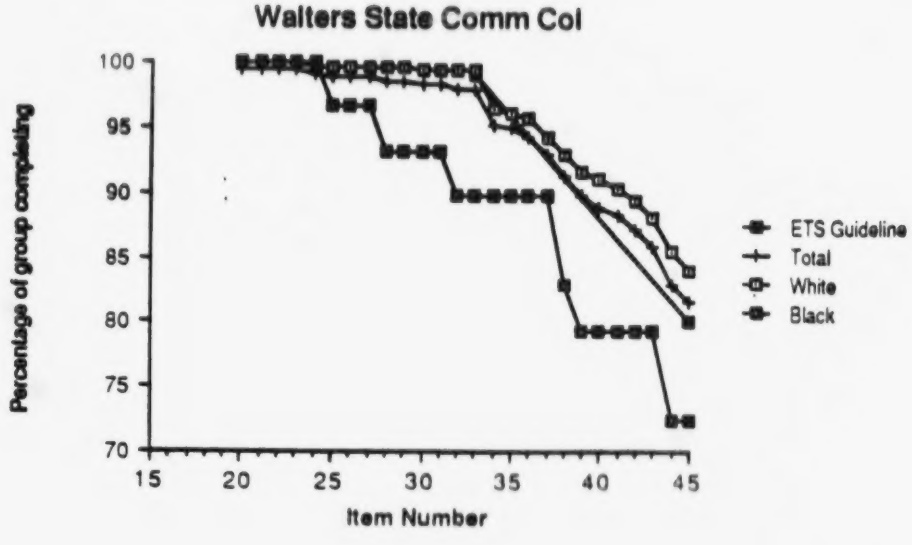


Figure 20

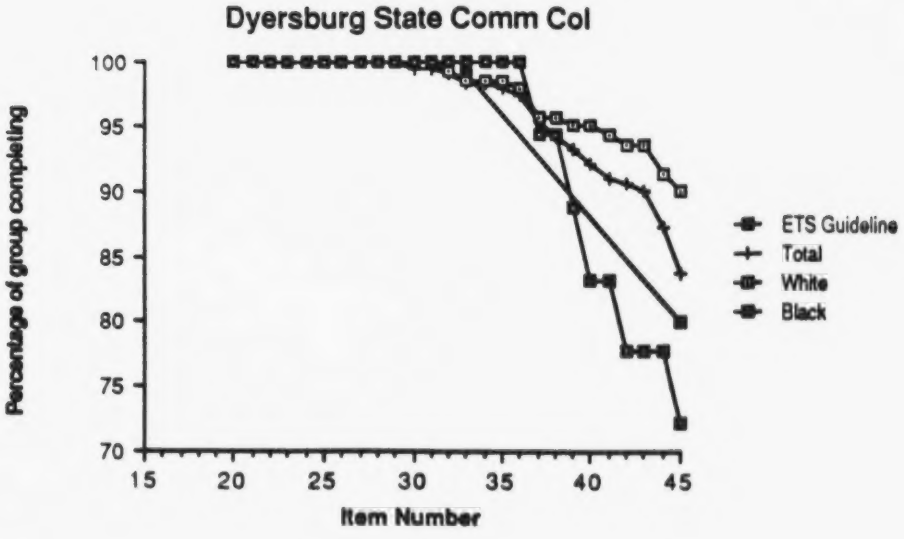
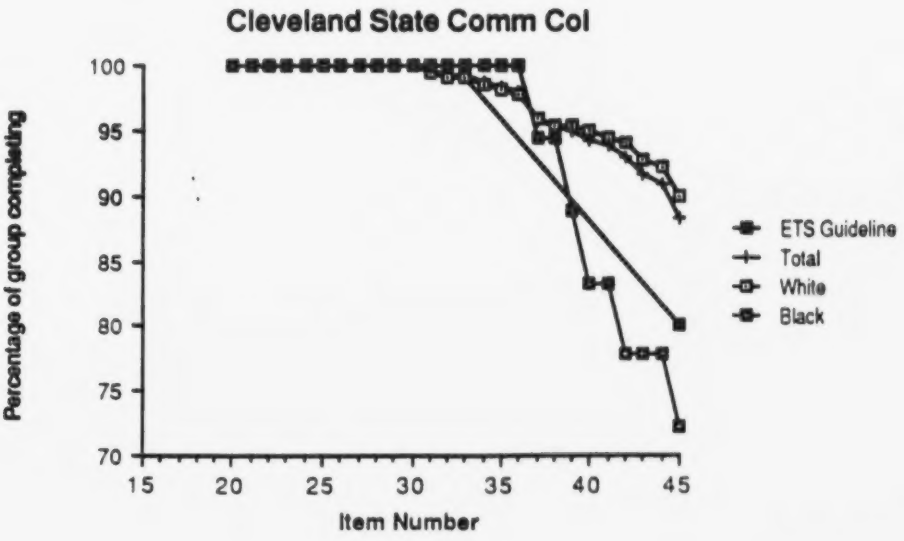


Figure 21



At Shelby State (Figure 15) 507 blacks and 142 whites were tested, with a black completion rate of 60.4 percent and a white rate of 83.1 percent. Volunteer State (Figure 16) tested only 16 blacks and their completion rate was 50 percent, as compared to a completion rate of 84.7 percent for the 307 whites tested. The completion rate at Chattanooga State (Figure 17) for the 83 blacks and 385 whites tested were 55.4 percent and 84.7 percent respectively. At Roane State (Figure 18) 28 blacks and 395 whites took the AAPP, with a black completion rate of 75 percent and a white completion rate of 88.6 percent. The completion rate for the 29 black students and 311 white students at Walters State (Figure 19) was 72.4 percent and 83.9 percent respectively. At Dyersburg State (Figure 20) the black completion rate for the 46 tested was 63 percent, while the rate for the 128 tested whites was 90.1 percent. Finally at Cleveland State (Figure 21) only 18 blacks were tested and of those 72.2 percent reached the last item. For whites at Cleveland State, of the 219 tested, 90 percent were able to complete the test.

Speededness for the technical institutions. Four technical institutes are operated by the State Board of Regents. These schools are State Tech at Nashville, State Tech at Memphis, State Tech at Knoxville and Tricities State Tech. The results of the analysis of the speededness of the AAPP parallels that of the community colleges. At all four of the technical institutes, white students' completion rates were higher than those of blacks. At two of the schools, State Tech at Memphis (Figure 22) and State Tech at Nashville (Figure 23), the differences in completion rates were profound.

Figure 22

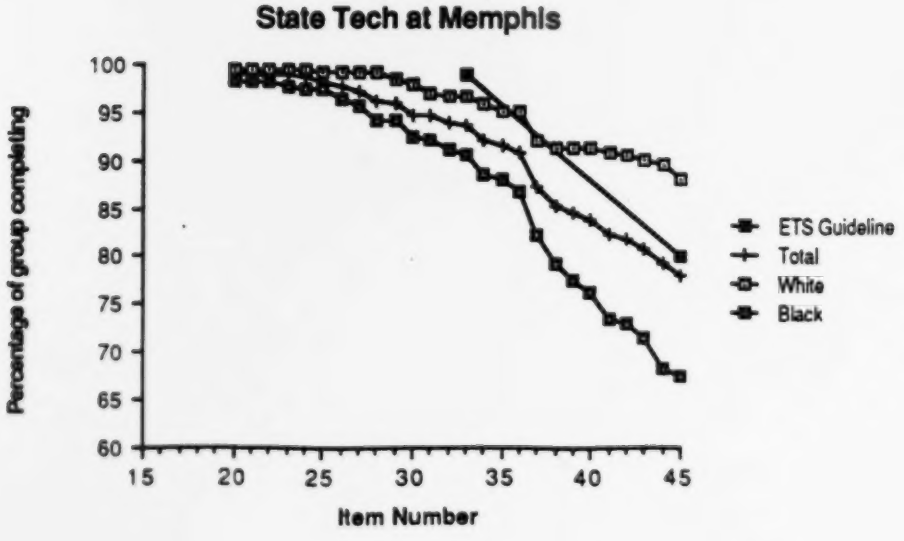
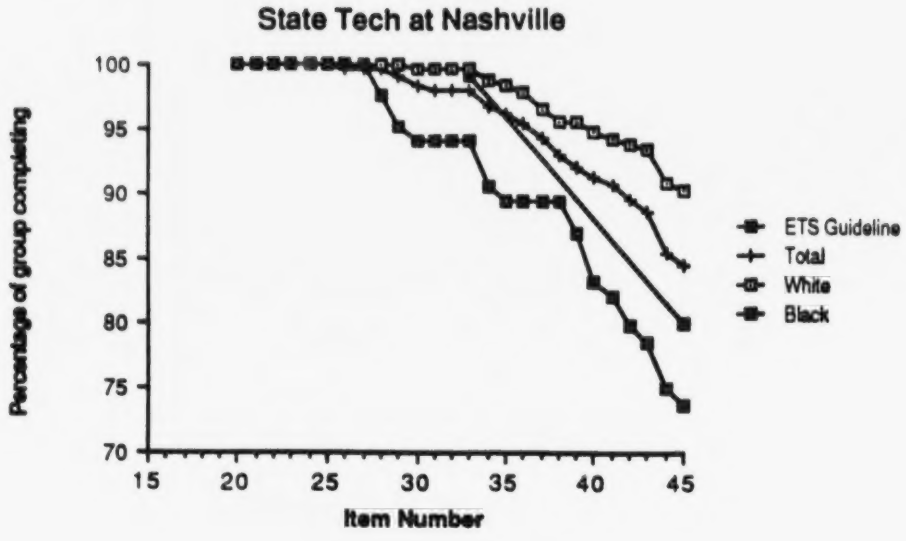


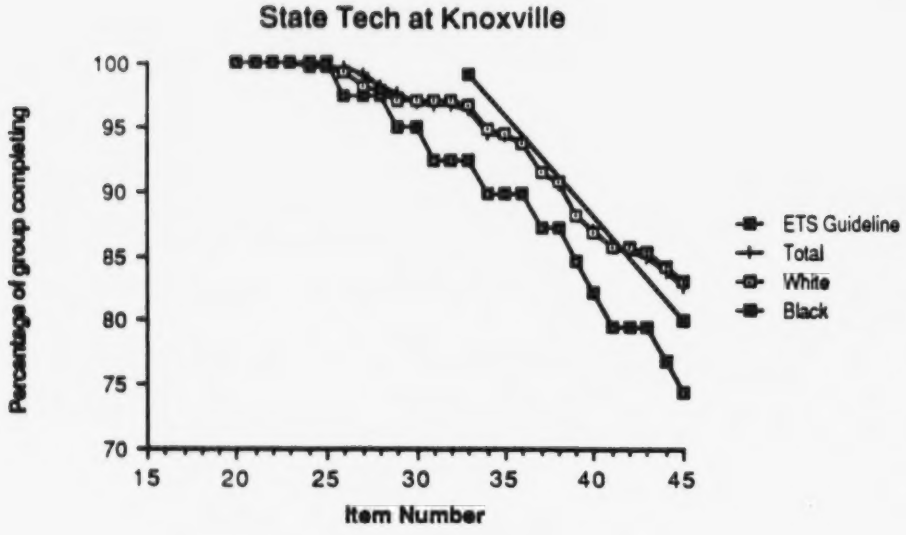
Figure 23



At Memphis (Figure 22), the completion rate for the 244 white students taking the test was 88.1 percent. In contrast, among the 228 black students, only 67.5 percent reached the last item. While virtually all whites at Memphis were able to reach at least 75 percent of the items on the AAPP, only 91 percent of the blacks were so able. At Nashville (Figure 23) only 73.3 percent of the 84 blacks completed the exam compared with 90.4 percent of the 229 whites who took the test.

Figure 24 displays the pattern of completion for whites and blacks at State Tech at Knoxville.

Figure 24

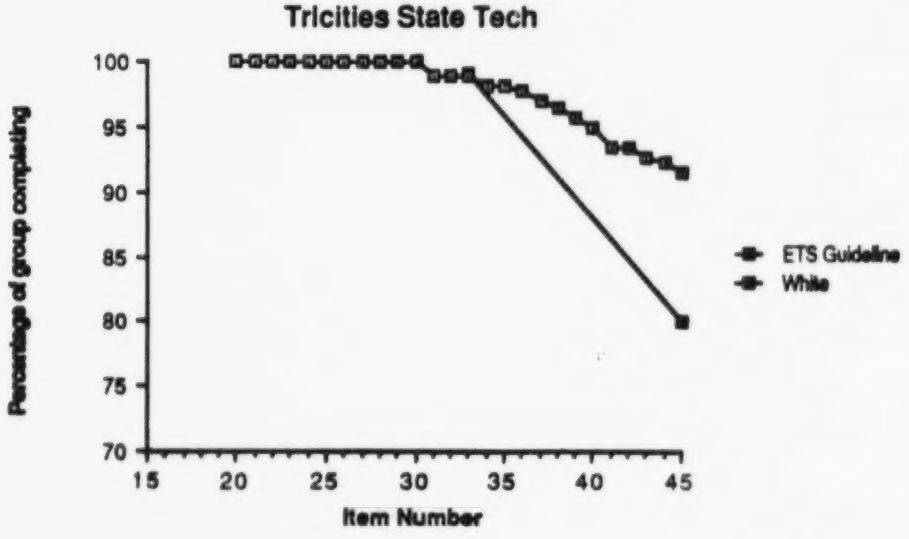


The differences between 266 whites and 39 blacks are relatively small. This is largely due to the fact that neither group's completion rate was very high. Among whites, only 83.1 percent of test takers completed the exam, as did 74.4 percent of blacks. Thus, among blacks at Knoxville, the AAPP clearly contains a speeded component, while for whites the component appears to be minimally present.

At Tricities State Tech 276 whites and 6 blacks were examined. As so few blacks are represented, it was felt that conclusions about black completion rates at Tricities would not be stable over time and thus were

not warranted. The data displayed in Figure 25 represents the completion curve for white students only at Tricities.

Figure 25



As Figure 25 suggests, the completion rate for white students is quite high with 91.7 percent of white students completing the exam within the allotted time.

Conclusions and Recommendations

In responding to the SBR's request to investigate the impact of time limits on the performance of students taking the AAPP reading comprehension test, the question was framed as follows: Is the test a speeded test or a timed-power test? The question was addressed by

examining two criteria for a test to be considered timed-power as opposed to speed, the proportion of persons finishing the test and the proportion finishing at least 75 percent of the items.

It is important to note that this present study does not attempt to estimate how much impact speed requirements have on actual test scores. This would require calculation of the work rate as well as the efficiency with which items are answered correctly. This present study was designed to address the question of whether or not the established time limit was a factor in the test results, rather than an attempt to predict what individuals would do with additional allotted time. Nor does this study address the important related issue of the validity of the instrument as a placement test.

Based on our analysis of the Fall 1986 AAPP pretest data, we have demonstrated differential completion rates for blacks and whites at the SBR institutions. The trend observed was for whites to have higher completion rates than blacks. This trend was noted at all institutions examined. At some institutions, especially among most of the community colleges and technical institutes, the differences between blacks and whites

are so profound that, when judging by the criteria for speededness developed by ETS, one may conclude that, for blacks only, a speed component is a significant factor in determining final scores and, thus, placement into remedial and developmental courses. For whites only, the AAPP functions as indicated in the test manual, namely as a timed-power test in which final scores are largely attributable to the ability to read with comprehension. These findings are consistent with those reported earlier by Flores and Seaman (1978).

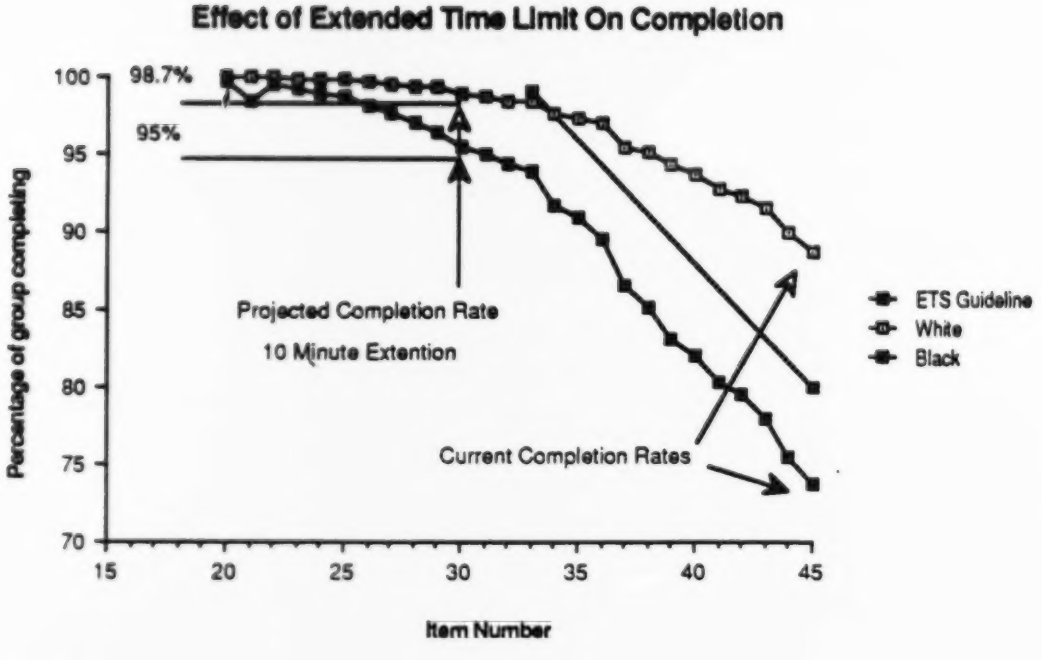
To the extent that success in college level courses can be attributed to speed, as distinct from understanding of what has been read, then a speed factor in a placement test may be desirable. At this time no such demonstration has been made. Indeed, the curricular thrust of the remedial and developmental courses is in terms of reading accuracy at the expense, if necessary, of reading rate. Thus the differential performance of blacks and whites on the AAPP attributable to speed may be said to disproportionally penalize black students when, in fact, speed has not been demonstrated to be an important factor in success in college level work. The inclusion of a component of variablility attributable to speed, for blacks, adds an

element of psychometric noise in the determination of an accurate level of reading comprehension for all students and especially black students.

As a timed-power test involves concerns for both speed of responding and the difficulty of items the "Comfortable time limit" (Nunnally, 1978, p. 632), or the amount of time required for 90 percent of the persons taking the test to complete it, is often employed with timed-power tests . Research indicates that increasing or decreasing time as much as 100 percent and 50 percent, respectively, of the comfortable time will not seriously affect any of the psychometric properties of the test or appreciably change most students' scores. However, the mean score of a group may increase with extended time limits (Nunnally, 1978, p. 633). For some tests, continuing to stare at items for which one does not know the answer does not increase one's scores. If one assumes a consistent work rate through the test (45 items divided by 30 minutes yields 40 seconds per item) of 40 seconds per item (Donlon, 1973), then an increase of 10 minutes in the time allotted for the test might enable approximately 95 percent of black test takers and 98 percent of white test takers, on the average, to reach all items.

This figure was derived directly from the current distribution of completion rates for whites and blacks. For example, from the current data we observe that 73.7 percent of blacks reach the last item and 95 percent of blacks reach item 30 within the allotted 30 minutes. Now assuming a consistent work rate of 40 seconds per item and adding an additional 10 minutes of time, an additional 15 items might thus be reached by 95 percent of current black test takers. Figure 26 graphically illustrates this solution.

Figure 26



This extension would be within the time limit increases which are psychometrically acceptable, as indicated by Nunnally, and well within Nunnally's 90 percent "comfortable time" completion rate for timed-power tests for both blacks and whites.

The addition of 10 minutes to the time allotted for the AAPP reading comprehension subtest would allow for the substantial reduction of the speed component for blacks, thus focusing placement decisions on reading comprehension for both blacks and whites rather than differentially adding a speed component for blacks. This reduction of the speed component is desirable. As Schwarz (1971) notes, time may introduce extraneous sources of variation into the measurement of performance that may be unwanted if the time factor is unrelated to criterion performance, which, in this case, is reading comprehension for college level material.

For a speed component to be of use in a placement test, it must be hypothesized that speed of performance provides an index of the <u>overall facility</u> of the test taker to handle more advanced and abstract work. For this argument to be advanced, it must be shown that agility with basic reading passages is indeed a good predictor of one's ability to handle more complex

material. At this time evidence supporting this hypothesis is not available. These recommended increases in the time allotted for the AAPP reading comprehension subtest are in keeping with the New Jersey experience in which the time-to-complete a revised 50-item version of the AAPP Reading Comprehension Test was increased to 50 minutes to insure a 90 percent completion rate.

While the time extension recommended here should enable test takers to respond to more of the items, and thus, have a completion rate compatible with that of a timed-power test, it would not assure any higher total score. Furthermore, it is not warranted to assume that reducing the speed component would make the test a more, nor less, valid instrument upon which to base placement decisions.

We add an opinion about the reading posttest using Form B, since it is a concern expressed by some institutional directors of developmental studies. We understand that the posttest is given after the student has passed the developmental course and is being moved into college level courses. Unlike the pretest given for placement purposes, the posttest results have no effect on the student and there is no incentive for the

student to do well. It is not surprising to find students, whose reading has improved with developmental instruction, nonetheless score lower on the second test, because there is nothing to be lost, nor gained, thereby. With such different conditions attached to the pretest and posttest, we can see no basis for comparing their results.

References

Academic preparation for college: What students need to know and be able to do. (1983). New York, NY: College Board Publications.

Cronbach, L. J., & Warrington, W. G. (1951). Time limit tests: Estimating their reliability and degree of speeding. Psychometrika, 14, 167-188.

Donlon, T. P. (1973, November). Establishing appropriate time limits for tests. Paper presented at the meeting of the Northeast Educational Research Association, Ellenville, NY. (ERIC Document Reproduction Service No. ED 118 590)

Donlon, T. P. (1975, April). Estimating rate-of-work parameters in the assessment of test speededness. Paper presented at the meeting of the National Council on Measurement in Education, Washington, D. C. (ERIC Document Reproduction Service No. ED 115 671)

Donlon, T. P. (1980a). An annotated bibliography of studies of test speededness. (Report No. GREB-76-9R). Princeton, NJ: Educational Testing Service. (ERIC Document Reproduction Service No. ED 241 565).

Donlon, T. F. (1980b). An exploratory study of the implications of test speededness. (Report No. GREB-76-9P). Princeton, NJ: Educational Testing Service. ERIC Document Reproduction Service No. 241 560).

Flores, T. B., & Seaman, D. F. (1978). A comparative study of adult student performance on timed GED tests in Texas. (ERIC Document Reproduction Service No. ED 154 138)

Gallagher, P., & Others. (1985). Time as a factor in test performance in a state mandated testing program. (ERIC Document Reproduction Service No. ED 275 729).

Gulliksen, H. (1950). Theory of mental tests. New York: John Wiley & Sons.

Immerman, M. A. (1980). The effect of eliminating time restraints on a standardized test with American Indian adults. (ERIC Document Reproduction Service Report No. ED 196 584)

Jolly, S. J. (1985, March). The effect of test speededness and random guessing on the validity of reading comprehension scores. Paper presented at the meeting of the American Educational Research Association, Chicago, IL. (ERIC Document Reproduction Service Report No. ED 262 063)

Nunnally, J. C. (1978). Psychometric Means. New York: McGraw-Hill.

Reilly, R. R., & Evans, F. R. (1974, September). The effects of time test limits on performance of culturally defined groups. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA. (ERIC Document Reproduction Service Report No. ED 102 276).

Stafford, R. E. (1971). The speededness quotient: A new descriptive statistic for tests. Journal of Educational Measurement, 8, 275-278.

State Board of Regents. (1986). Analysis of enrollment of first-time freshmen under 21 years of age by all Tennessee high schools and all SBR institutions. (Report No. 1) Nashville, TN.

Schwarz, P. A. (1971). Prediction instruments for educational outcomes. In R. L. Thorndike (Ed.), Educational measurement (pp. 303-331). Washington, DC: American Council on Education.